



Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men

Suvitaival, Tommi; Bondia-pons, Isabel; Yetukuri, Laxman; Pöhö, Päivi; Nolan, John J.; Hyötyläinen, Tuulia; Kuusisto, Johanna; Oreši, Matej

Published in:
Metabolism

DOI:
[10.1016/j.metabol.2017.08.014](https://doi.org/10.1016/j.metabol.2017.08.014)

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY-NC-ND](#)

Citation for published version (APA):
Suvitaival, T., Bondia-pons, I., Yetukuri, L., Pöhö, P., Nolan, J. J., Hyötyläinen, T., Kuusisto, J., & Oreši, M. (2018). Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men. *Metabolism*, 78, 1-12. <https://doi.org/10.1016/j.metabol.2017.08.014>

Available online at www.sciencedirect.com

Metabolism

www.metabolismjournal.com

Clinical Science

Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men



Tommi Suviataival^{a,*}, Isabel Bondia-Pons^a, Laxman Yetukuri^{b,c}, Päivi Pöhö^{c,d}, John J. Nolan^a, Tuulia Hyötyläinen^{a,e,f}, Johanna Kuusisto^{g,1}, Matej Orešič^{a,e,h,1}

^a Steno Diabetes Center Copenhagen, DK-2820 Gentofte, Denmark^b Institute for Molecular Medicine Finland, FI-00014, University of Helsinki, Helsinki, Finland^c VTT Technical Research Centre of Finland, FI-02044 VTT, Espoo, Finland^d Faculty of Pharmacy, FI-00014, University of Helsinki, Helsinki, Finland^e Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, FI-20520 Turku, Finland^f Department of Chemistry, Örebro University, 702 81 Örebro, Sweden^g Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, FI-70211 Kuopio, Finland^h School of Medical Sciences, Örebro University, 702 81 Örebro, Sweden

ARTICLE INFO

Article history:

Received 24 November 2016

Accepted 26 August 2017

Keywords:

Lipidomics

Mass-spectrometry

METSIM study

Plasma profiling

Type 2 diabetes mellitus

ABSTRACT

Background. There is a need for early markers to track and predict the development of type 2 diabetes mellitus (T2DM) from the state of normal glucose tolerance through prediabetes. In this study we tested whether the plasma molecular lipidome has biomarker potential to predicting the onset of T2DM.

Methods. We applied global lipidomic profiling on plasma samples from well-phenotyped men (107 cases, 216 controls) participating in the longitudinal METSIM study at baseline and at five-year follow-up. To validate the lipid markers, an additional study with a representative sample of adult male population (n = 631) was also conducted. A total of 277 plasma lipids were analyzed using the lipidomics platform based on ultra-performance liquid chromatography coupled to time-of-flight mass spectrometry. Lipids with the highest predictive power for the development of T2DM were computationally selected, validated and compared to standard risk models without lipids.

Results. A persistent lipid signature with higher levels of triacylglycerols and diacyl-phospholipids as well as lower levels of alkylacyl phosphatidylcholines was observed in progressors to T2DM. Lysophosphatidylcholine acyl C18:2 (LysoPC(18:2)), phosphatidylcholines PC(32:1), PC(34:2e) and PC(36:1), and triacylglycerol TG(17:1/18:1/18:2) were selected to the full model that included metabolic risk factors and FINDRISC variables. When further adjusting for BMI and age,

Abbreviations: 2HPG, 2-hour plasma glucose; ALT, alanine aminotransferase; AUC, area under curve; Cer, ceramide; ChoE, cholesterol ester; FPG, fasting plasma glucose; FINDRISC, Finnish diabetes risk score; hs-CRP, high-sensitivity C-reactive protein; IDI, integrated discrimination improvement; IFG, impaired fasting glucose; IGT, impaired glucose tolerance; LysoPC, lysophosphatidylcholine; LysoPE, lysophosphatidylethanolamine; NAFLD, non-alcoholic fatty liver disease; NGT, normal glucose tolerance; OGTT, oral glucose tolerance test; PC, diacyl phosphatidylcholine; PCe, alkylacyl phosphatidylcholine; PE, phosphatidylethanolamine; PEe, alkylacyl phosphatidylethanolamine; PL, glycerophospholipid; ROC, receiver operating characteristic; SM, sphingomyelin; TG, triacylglycerol; UPLC-QTOFMS, ultra-performance liquid chromatography coupled to time-of-flight mass spectrometry.

* Corresponding author at: Steno Diabetes Center Copenhagen, Niels Steensens Vej 2-4, DK-2820 Gentofte, Denmark.

E-mail address: tommi.raimo.leo.suviataival@regionh.dk (T. Suviataival).

¹ Shared senior authorship.

<https://doi.org/10.1016/j.metabol.2017.08.014>

0026-0495/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

these lipids had respective odds ratios of 0.32, 2.4, 0.50, 2.2 and 0.31 (all $p < 0.05$) for progression to T2DM. The independently-validated predictive power improved in all pairwise comparisons between the lipid model and the respective standard risk model without the lipids (integrated discrimination improvement IDI > 0 ; $p < 0.05$). Notably, the lipid models remained predictive of the development of T2DM in the fasting plasma glucose-matched subset of the validation study.

Conclusion. This study indicates that a lipid signature characteristic of T2DM is present years before the diagnosis and improves prediction of progression to T2DM. Molecular lipid biomarkers were shown to have predictive power also in a high-risk group, where standard risk factors are not helpful at distinguishing progressors from non-progressors.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Type 2 diabetes mellitus (T2DM) is a multifactorial metabolic disorder characterized by hyperglycemia, which results from impaired insulin secretion of pancreatic β -cells and from ineffective cellular response to insulin [1]. Diabetes is preceded by a prodromal phase that may last many years. Prediabetes is currently characterized, once glucose has become elevated, by impaired fasting glucose (IFG), impaired glucose tolerance (IGT) or both. IFG and IGT are not equivalent by metabolic terms and, most likely, reflect different pathophysiological states leading to T2DM [2]. Consequently, they may require different treatments.

The onset of T2DM in prediabetic individuals can be prevented—or at least delayed—by pharmacological or life-style interventions [3]. Ideally, knowledge about the underlying pathophysiological characteristics associated with either fasting or postprandial glucose dysregulation could be helpful at optimizing the efficacy of treatments [4]. Therefore, there is a need for predictive tools enabling the efficient and accurate tracking of progression from a state of normal glucose tolerance (NGT) to pre-diabetes (IFG, IGT or both) and finally to T2DM. So far, several T2DM risk models and scores have been developed as prognostic tools. These models are mainly based on established risk factors of T2DM and lack the specificity needed for the use in clinical practice [5–7]. A comprehensive review [7] of almost one hundred T2DM predictive models showed that the most typical non-lipid risk factors included in a T2DM risk prediction model are the age, obesity, gender, family history, blood pressure, FPG, 2-h postprandial plasma glucose and glycated hemoglobin (HbA1c). The most typical lipid-based risk factors are the HDL-C, LDL-C and triglycerides (TG).

Several metabolites have indeed been found to be associated with dysglycemia and T2DM [8]. Fasting plasma levels of branched-chain amino acids and aromatic amino acids have been reported to be associated with the increased risk of T2DM in prospective studies [9,10]. Population-based studies have revealed that glycine, lysophosphatidylcholine acyl C18:2 (LysoPC(18:2)), acetylcarnitine [11], α -hydroxybutyrate [12] as well as specific choline-containing phospholipids [13] are predictive of abnormalities in glucose tolerance. Recently, a novel IGT test with metabolite markers was reported [14,15].

Knowledge about global changes in the plasma lipidome during the progression from NGT or prediabetes to T2DM still is limited. The potential of plasma lipids as predictors of T2DM has not been thoroughly validated. A recent targeted

lipidomic profiling from the ADVANCE study demonstrated [16] the potential of circulating molecular lipids as biomarkers of cardiovascular risk in T2DM. Specific lipids such as TGs with low carbon number and low double bond content, which also characterize increased liver fat content [17], have been associated with insulin resistance and elevated risk of T2DM [18,19]. Specific plasma ceramides, which are elevated in obese subjects with T2DM, have also been found [20] to reflect the severity of insulin resistance. Moreover, ceramides present in LDL have been found [21] to be elevated in T2DM, promoting inflammation and skeletal muscle insulin resistance.

The plasma lipidome is sensitive to the amount of liver fat and has been shown [17] to predict non-alcoholic fatty liver disease (NAFLD). Since liver fat is a risk factor of T2DM [22], we hypothesize that aberrations in the levels of plasma molecular lipids could also be independent predictors of T2DM. To test this hypothesis, we applied a global lipidomic profiling approach to a prospective study of well-phenotyped male subjects with an aim (a) to identify a lipid signature that precedes T2DM, and (b) to build and validate a lipid-based model that can predict progression to T2DM in adult men.

2. Methods

2.1. Participants

A total of 10,197 men participated in the population-based METSIM (METabolic Syndrome In Men) Study [23]. The 5-year follow-up data were available for 3408 men. Glucose tolerance status was assessed according to the ADA criteria [23] on plasma glucose levels in OGTT and HbA1c measurements. In total, 2916 men were non-diabetic at baseline both by the OGTT and HbA1c criteria (FPG < 7.0 mmol/L and 2HPG < 11.1 mmol/L and HbA1c $< 6.5\%$). The study protocol was approved by the Ethics Committee of the University of Eastern Finland and Kuopio University Hospital. The study was conducted in accordance with the principles of the Helsinki Declaration. Each participant provided a written informed consent.

2.2. Study Design

Participants were divided into a discovery set ($n = 323$) and a validation set ($n = 631$) to fit and validate the models, respectively. The baseline clinical characteristics for both groups are shown in Table 1.

Table 1 – Baseline biochemical measurements and clinical characteristics of the METSIM study participants (n = 958) selected for the discovery set (n = 323) and for the validation set (n = 631).

| | Discovery set | | | | Validation set | | | |
|----------------------------------|-----------------|--------------|--------------|---------------|-----------------|--------------|--------------|---------------|
| | Non progressors | | | Progressors | Non progressors | | | Progressors |
| | NGT | Pre-diabetic | All | All | NGT | Pre-diabetic | All | All |
| n | 129 | 87 | 216 | 107 | 127 | 481 | 608 | 23 |
| Age (years) | 59.3 ± 5.1 | 60.8 ± 5.4 | 59.9 ± 5.3 | 59.0 ± 5.7 | 52.2 ± 5.0 | 53.9 ± 4.4 | 53.6 ± 5.0 | 54.2 ± 4.2 |
| Family history of diabetes (%) | 44.2 | 46.0 | 44.9 | 59.8* | 48.8 | 52.2 | 51.5 | 52.2 |
| BMI (kg/m ²) | 26.1 ± 2.4 | 26.2 ± 2.1 | 26.2 ± 2.3 | 28.6 ± 3.8* | 25.1 ± 2.8 | 26.4 ± 3.1 | 26.1 ± 3.2 | 29.4 ± 4.2* |
| Waist circumference (cm) | 95.1 ± 7.2 | 95.6 ± 6.5 | 95.3 ± 6.9 | 102.4 ± 10.4* | 98.3 ± 5.9 | 100.3 ± 4.7 | 99.9 ± 5.9 | 103.7 ± 8.1* |
| Systolic blood pressure (mm Hg) | 132.1 ± 15.1 | 133.9 ± 15.3 | 132.8 ± 15.2 | 143.0 ± 16.0* | 129.1 ± 12.5 | 133.8 ± 14.1 | 132.8 ± 14.0 | 143.6 ± 17.7* |
| Diastolic blood pressure (mm Hg) | 84.3 ± 8.6 | 85.1 ± 9.8 | 84.6 ± 9.1 | 89.8 ± 8.6* | 84.1 ± 7.7 | 87.3 ± 8.5 | 86.6 ± 8.4 | 94.2 ± 9.9* |
| Fasting plasma glucose (mmol/l) | 5.2 ± 0.2 | 5.1 ± 0.2 | 5.2 ± 0.2 | 6.0 ± 0.5* | 5.3 ± 0.2 | 5.8 ± 0.2 | 5.7 ± 0.4 | 6.2 ± 0.5* |
| 2-hour plasma glucose (mmol/l) | 5.0 ± 0.8 | 5.1 ± 1.0 | 5.0 ± 0.9 | 8.1 ± 2.0* | 5.0 ± 1.2 | 5.8 ± 1.1 | 5.6 ± 1.5 | 8.2 ± 1.8* |
| Fasting plasma insulin (mU/l) | 5.3 ± 2.8 | 5.7 ± 2.7 | 5.5 ± 2.8 | 11.9 ± 7.0* | 5.6 ± 2.9 | 7.6 ± 3.6 | 7.2 ± 4.5 | 12.3 ± 5.8* |
| 2-hour plasma insulin (mU/l) | 28.7 ± 25.5 | 31.1 ± 20.7 | 29.7 ± 23.6 | 100.3 ± 68.4* | 32.0 ± 31.2 | 41.3 ± 26.6 | 39.4 ± 27.8 | 97.6 ± 64.0* |
| Hemoglobin A1C (%) | 5.5 ± 0.2 | 5.8 ± 0.2 | 5.6 ± 0.3 | 6.3 ± 0.8* | 5.4 ± 0.2 | 5.8 ± 0.1 | 5.6 ± 0.3 | 5.9 ± 0.4* |
| LDL cholesterol (mmol/l) | 3.4 ± 0.9 | 3.3 ± 0.8 | 3.4 ± 0.8 | 3.3 ± 0.8 | 3.6 ± 0.8 | 3.5 ± 0.7 | 3.5 ± 0.8 | 3.5 ± 1.0 |
| HDL cholesterol (mmol/l) | 1.5 ± 0.4 | 1.6 ± 0.4 | 1.5 ± 0.4 | 1.3 ± 0.4 | 1.5 ± 0.4 | 1.5 ± 0.4 | 1.5 ± 0.4 | 1.4 ± 0.4 |
| Total triacylglycerols (mmol/l) | 1.1 ± 0.5 | 1.1 ± 0.5 | 1.1 ± 0.5 | 1.9 ± 1.2* | 1.3 ± 0.7 | 1.4 ± 0.9 | 1.4 ± 0.8 | 1.8 ± 0.9* |
| Fat mass (%) | 23.2 ± 6.4 | 24.5 ± 6.3 | 23.7 ± 6.4 | 25.5 ± 6.1 | 18.9 ± 3.8 | 20.4 ± 4.1 | 20.0 ± 4.4 | 24.1 ± 4.5* |
| Serum ALT (U/l) | 24.6 ± 12.5 | 24.9 ± 10.5 | 24.8 ± 11.7 | 40.0 ± 25.4* | 27.5 ± 15.5 | 31.2 ± 18.5 | 30.4 ± 18.5 | 41.8 ± 18.5* |
| Plasma hs-CRP (mg/l) | 1.7 ± 2.6 | 1.7 ± 2.5 | 1.7 ± 2.3 | 2.2 ± 2.4 | 1.1 ± 2.8 | 1.8 ± 1.3 | 1.7 ± 1.3 | 2.3 ± 1.3 |

Data are presented in mean ± SD for continuous variable or percentages for categorical variables. ANOVA followed by post hoc Tukey comparisons for continuous variables and Chi-squared test (χ^2) for categorical variables ($p < 0.001$), *non-progressor significantly different from progressor to type 2 diabetes.

The discovery set comprised a case/control design with 107 cases (i.e., progressing individuals who were non-diabetic at baseline but developed T2DM during the five-year follow-up), and 216 age-matched controls (i.e., non-progressing individuals who were non-diabetic at baseline and did not develop T2DM during the five-year follow-up). The criteria for a new-onset T2DM were FPG ≥ 7.0 mmol/L and/or 2HPG ≥ 11.1 mmol/L and/or HbA1c $\geq 6.5\%$, and/or a physician-based diagnosis and anti-diabetic medication started between the baseline and follow-up examinations. Participants with incident T2DM, diagnosis of T2DM between baseline and follow-up ($n = 54$) or greatest glucose AUC at follow-up ($n = 53$ of the cases diagnosed with T2DM at follow-up) were included as T2DM progressors for the present study. A total of 504 participants were normoglycemic (FPG < 5.6 mmol/L, 2HPG < 7.8 mmol/L and HbA1c $< 6.5\%$) both at the baseline and follow-up examinations. Out of these, 216 age-matched subjects from the top-end of the body mass index (BMI) distribution and from the lower end of the distribution of glucose area under the glucose curve (AUC) in an OGTT at follow-up were included as T2DM non-progressors for the discovery set. There were no cases of IFG or IGT among the non-progressors.

Non-diabetic subjects (by both the OGTT and HbA1c criteria at baseline) from the remaining 1593 participants of the METSIM Study were randomly selected to the validation set. Clinical data at baseline and follow-up were available for 631 subjects. T2DM

non-progressors in the validation set were characterized by a broad distribution of glucose tolerance with 21% NGT and 79% pre-diabetic subjects (77% IFG; 2% IGT; 7% IFG + IGT; 14% with $5.7 < \text{HbA1c}(\%) < 6.4$; Supplementary Fig. S1). T2DM prediction in the validation set was tested in the full set, in the FPG-matched set of non-progressors and progressors, and in a set including only progressors and prediabetic non-progressors.

An unfrozen plasma sample (EDTA) was available from all participants at baseline. An additional plasma sample at follow-up was only available from the participants in the discovery set. Plasma was separated by centrifugation ($1300 \times g$; 15 min; $+4^\circ\text{C}$), frozen immediately at -80°C , and stored until analysis.

2.3. Measurements

Clinical and biochemical measurements were performed as described in the Supplementary Methods S2.3.1 and earlier [23]. Molecular serum lipids were analyzed at VTT Technical Research Centre of Finland (Espoo, Finland) with a methodology described in the Supplementary Methods S2.3.2 and earlier [24].

2.4. Statistical Analysis

Clinical characteristics of the study participants are presented as mean ± standard deviation (SD) for continuous variables

and as percentages for categorical variables. The p-values for continuous and categorical variables were computed using Student's t-test and chi-squared test, respectively.

Lipidomic data were base-2 log-transformed, and scaled to zero mean and unit variance. Gaussian mixture model (mclust2 package [25] for R) was fit on the data from the discovery set at baseline to identify coherent clusters of lipids. The number of clusters was selected based on the Bayesian information criterion (BIC) of the model.

Linear mixed-effect models (lme4 package [26] for R) were fit to the clustered discovery set at baseline to identify coherent early-time differences between progressors and non-progressors. Each model was adjusted for BMI and age, and included one of the lipid clusters as the dependent variable, the group label (progressor/non-progressor) as the fixed effect. Overall group differences were tested using F-statistic and post-hoc analyses were done using Tukey's all-groups comparisons.

Average profiles of the lipid clusters were compared between the groups after scaling all the data based on the lipid-specific mean and standard deviation of the baseline NGT non-progressors from the discovery set. Paired t-test was used to compare cluster profiles between the baseline and five-year follow-up time points in the discovery set. Non-paired t-test was used for comparing baseline cluster profiles between progressors and non-progressors in the discovery and validation sets (p-value < 0.01).

2.5. Predictive Modeling of Progression to Type 2 Diabetes

Logistic regression models were fit with the aim of predicting the progression to T2DM (i.e., the progressor/non-progressor label as the dependent variable of the model). Predictive lipids were selected in four ways: independently (the "Lipids" model), in combination with the Finnish Diabetes Risk Score variables ("Lipids and FINDRISC"), in combination with the clinical marker variables of metabolic syndrome (FPG, HDL-C, systolic blood pressure and total triacylglycerols; the "Lipids and MS" model), and in combination with both the FINDRISC and metabolic syndrome variables ("Lipids, FINDRISC and MS"). These four lipid models were compared to their respective counterparts without the lipid variables ("Random," "FINDRISC," "MS," and "FINDRISC and MS," respectively).

Lipid variables with the highest predictive power at the baseline measurement were forward-selected based on the area under the receiver operating characteristic curve (AUC) of the prediction of left-out samples in nested five-fold cross-validation [27] using the discovery set. Only lipid variables were selected; the FINDRISC and MS variables were always included in the models where their inclusion is stated. The FINDRISC model was fixed with the statistically significant coefficients reported in the original study [5].

Following the guideline [28] for the ratio between the limiting sample size and model complexity for a binary outcome variable, models with their top-five-selected lipid variables over the cross-validation were further tested with the independent validation set. The models were fit using the bootstrap-sampled [27] discovery set at baseline, and tested by predicting the validation set. Both for the cross-validated prediction of the discovery set and for the bootstrapped prediction of the validation set, the 95% confidence intervals for the AUC and the p-values for the test of difference

between the AUCs were computed based on the empirical distributions [29] over 200 bootstrap randomizations. Further, integrated discrimination improvement (IDI) [30] with 95% confidence intervals and the test of positive IDI between lipid and non-lipid models were computed in the same way.

The models were further tested with the validation set sub-sampled in two different ways: first, with FPG-matched balanced set of progressors and non-progressors ($n = 23$ and $n = 23$, respectively; Supplementary Table S1) in order to eliminate the effect of FPG as a confounding factor, and second, with progressors and prediabetic non-progressors only (i.e., high-risk subjects; $n = 23$ and $n = 481$, respectively). Difference between progressors and non-progressors was computed for each of the selected lipids in each of the data sets as well as in the high-risk subset and the FPG-matched subset of the validation set. Additionally, the difference was computed in a subset of the discovery set with a normal FPG at baseline ($n = 17$ progressors and matched non-progressors; Supplementary Table S2). Details of these sub-sampling experiments are explained in Supplementary Methods S2.5.

3. Results

3.1. Study Setting

Biochemical and clinical characteristics of subjects in the discovery set ($n = 323$) and the validation set ($n = 631$) are presented in Table 1. As expected, progressors were characterized by significantly higher levels of traditional risk factors of T2DM than the non-progressors.

A total of 277 molecular lipids were identified both in the discovery set and in the validation set. These lipids were glycerophospholipids (PL), such as diacyl phosphatidylcholines (PC), alkylacyl phosphatidylcholines (PCe), phosphatidylethanolamines (PE), alkylacyl phosphatidylethanolamines (PEe), lysophosphatidylcholines (LysoPC) and lysophosphatidylethanolamines (LysoPE), triacylglycerols (TG), cholesterol esters (ChoE), sphingomyelins (SM) and ceramides (Cer).

3.2. Baseline Lipid Profiles Associated With Progression to Type 2 Diabetes

The high degree of co-regulation in the lipidome was modeled with unsupervised probabilistic clustering fit to the baseline lipidomic data of the discovery set. The model identified 12 lipid clusters (LCs; Table 2), which to a large extent follow functional and structural groups (Supplementary Table S3).

Already five years before the diagnosis, marked differences in the lipidome were observed between progressors and non-progressors (Fig. 1). The lipidomic signature of progressors was characterized by elevated levels of four TG clusters (LC9 to LC12) and one PL cluster (LC8), and an opposite pattern of the ether PL (alkylacyl phospholipids) cluster (LC5). Several established risk markers of T2DM (BMI, waist circumference, HDL, LDL, total TG, and ALT levels) were positively correlated with the TG clusters (LCs 9–12), and negatively correlated with the ether PL cluster (LC5; Supplementary Fig. S2).

3.3. Lipidomic Profiles at the Five-Year Follow-Up Visit

The persistence of the early lipidomic signature characterizing progression to T2DM was assessed by comparing the

Table 2 – Summary of the lipid clusters, showing the cluster name, number of lipids in the cluster, lipid groups in the cluster, p-value of the difference between progressors and non-progressors in the discovery set at baseline, and example lipid species in the cluster.

| Cluster name | Cluster size | Cluster description | p-Value | Examples of lipid species |
|--------------|--------------|---|---------------------|---|
| LC1 | 25 | SM, Cer, PCs, PE, TG | 0.22 | SM(d18:1/16:1), PC(32:0), PC(34:2e), PE(38:2), TG(53:7) |
| LC2 | 4 | Cholesterol esters | 0.77 | ChoE(18:1), ChoE(18:2), ChoE(20:4), ChoE(20:5) |
| LC3 | 12 | LysoPC and LysoPE | 0.78 | LysoPC(14:0), LysoPC(18:2), LysoPE(18:0), LysoPE(18:2) |
| LC4 | 27 | PUFA-containing PC, PE, PCe, PEe, LysoPC | 0.34 | PC(18:0/22:6), PC(40:6e), PE(40:6), PE(40:7e), LysoPC(22:6) |
| LC5 | 30 | Alkylacyl (ether) phospholipids | 3×10^{-3} | PC(32:0e), PC(38:3e), PC(40:5e), PE(36:2e), PE(40:8e) |
| LC6 | 14 | Arachidonic acid containing- PC, PCe, and PEe | 0.05 | PC(16:0/20:4), PC(p18:0/20:4), PE(38:4e), PE(p18:0/20:4) |
| LC7 | 12 | Polyunsaturated long chain TG | 0.05 | TG(50:5), TG(52:6), TG(54:6), TG(54:7) |
| LC8 | 40 | PC and PE | 1×10^{-4} | PC(30:0), PC(40:5), PE(36:4), PE(40:6) |
| LC9 | 28 | Highly polyunsaturated long chain TG | 1×10^{-3} | TG(22:6/18:1/18:1), TG(56:7), TG(58:10), TG(62:12) |
| LC10 | 13 | Mono- and poly- unsaturated long chain TG | 2×10^{-6} | TG(50:3), TG(52:7), TG(58:7) |
| LC11 | 42 | Saturated and low unsaturated medium to long chain TG | 1×10^{-7} | TG(14:0/16:0/16:0), TG(16:0/18:0/12:0), TG(16:0/18:0/18:1) |
| LC12 | 30 | Abundant TG in plasma | 9×10^{-11} | TG(16:0/18:1/18:1), TG(55:5), TG(58:6) |

baseline lipid cluster profiles to the five-year follow-up profiles among the progressors, prediabetic non-progressors and NGT non-progressors of the discovery set ($n = 107$, 87 and 129, respectively). NGT non-progressors were similar to prediabetic non-progressors by their lipidomic profile. There were no significant differences between the groups either at baseline or at the five-year follow-up (Fig. 2). On the other hand, progressors had persistently higher levels of TGs (LC9–12), diacyl PLs (LC8), and lower levels of PCe lipids (LC5) when compared to NGT and pre-diabetic non-progressors. Specifically, the distinctive baseline signature observed in progressors remained unchanged at the five-year follow-up with no significant changes in the lipid cluster levels over the five-year period (Fig. 2). The only significant changes over the

five-year period were observed in non-progressors: All non-progressors had increased levels of arachidonic acid-containing PCs, NGT non-progressors had increased levels of highly-unsaturated long-chain TGs, and prediabetic non-progressors had increased levels of LysoPC, SM and ceramide-containing lipid clusters.

Previous studies [17–19] indicate that TGs with low double bond content and low carbon number are associated with liver fat, insulin resistance and T2DM. We also investigated the TG composition in the context of T2DM progression at baseline and follow-up. Indeed, we found that TGs with low double bond content and carbon number are more strongly associated with progression to T2DM as well as with overt T2DM (Fig. 3).

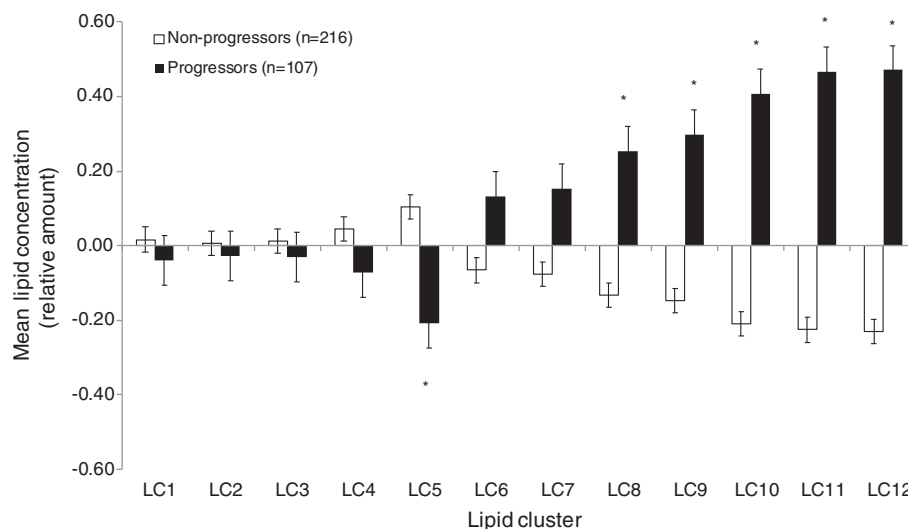


Fig. 1 – Mean lipid levels within each cluster, shown separately for non-progressors (white bars) and non-progressors to type 2 diabetes (black bars) in the discovery set. The data for each lipid are scaled to zero mean and unit variance. Average cluster level was calculated as the mean value of all scaled lipid levels belonging to the cluster. Statistical comparison was performed using linear mixed models (F-statistic; asterisk marker for $p < 0.01$). Error bars show the standard error of the mean. The clusters are summarized in Table 2.

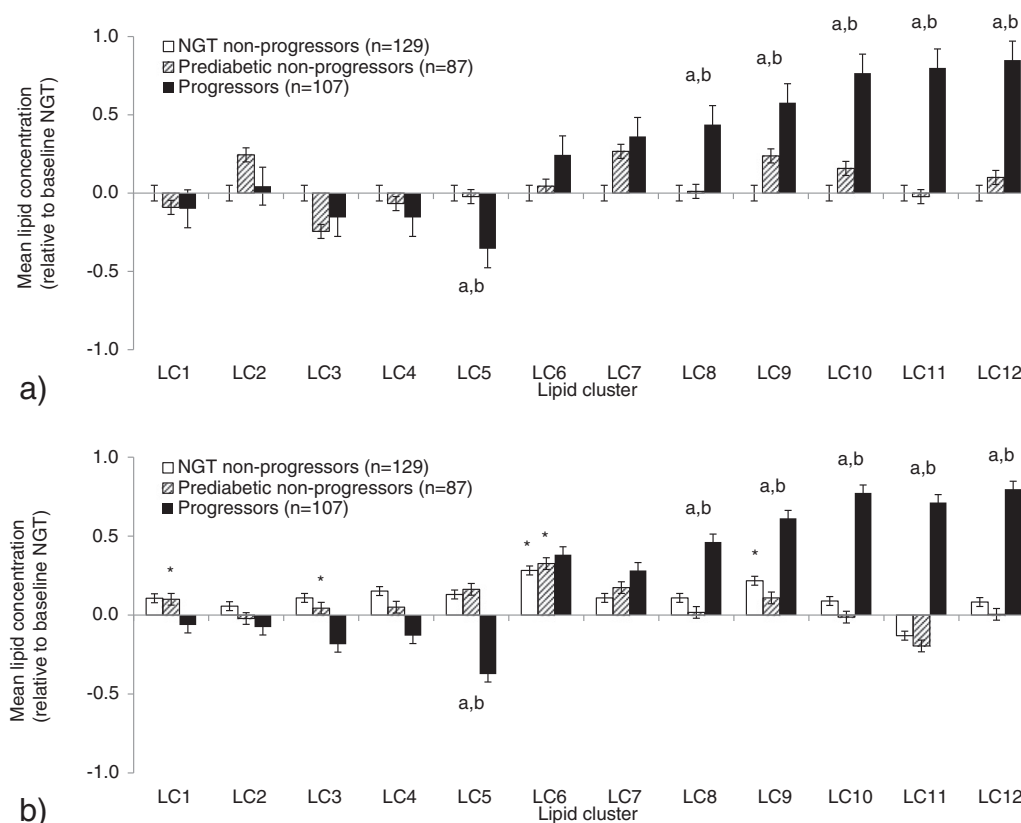


Fig. 2 – Mean levels of the lipid clusters for normal glucose-tolerant (NGT) non-progressors (white bars), prediabetic non-progressors (striped bars) and progressors to type 2 diabetes (black bars) in the discovery set (a) at baseline and (b) at five-year follow-up. For the simultaneous interpretation of group and time differences, the cluster levels are scaled based on the cluster-specific mean and standard deviation of the NGT non-progressor group at baseline, which is the reference group. The baseline and the five-year follow-up time points were compared with the paired two-sided t-test (asterisk marker for $p < 0.01$). The groups were compared within a time point with the non-paired two-sided t-test (marker “a” indicates a significant difference between progressors and NGT non-progressors; marker “b” indicates significant difference between progressors and prediabetic non-progressors; $p < 0.01$). Error bars show the standard error of the mean. The clusters are summarized in Table 2.

3.4. Lipid-Based Model to Predict Type 2 Diabetes

In total, 11 lipids were selected to the four prediction models (Table 3). The lipids that predicted progression to T2DM in the cross-validated discovery set were the LysoPC(18:2), three TGs (TG(50:1), TG(54:5), TG(56:4)) and the ether lipid PC(42:6e). The cross-validated prediction performance of the “Lipids” model, which consisted of these selected lipids only, was better than the random baseline (cross-validated area under the receiver operating characteristic curve, AUC “Lipids” = 0.842; AUC “Random” = 0.50; integrated discrimination improvement, IDI = 0.359; $p < 0.05$ for the test of positive IDI; Table 4; Supplementary Fig. S3).

When the FINDRISC variables were included as predictors, the selected lipids were the same as with the “Lipids” model except for the TG(51:5), which replaced the TG(56:4). The cross-validated prediction performance of the “Lipids and FINDRISC” model consisting of the selected lipids and the FINDRISC variables was better than the model with the FINDRISC variables only (AUC 0.864 vs. 0.755; IDI 0.282; $p < 0.05$).

When the metabolic syndrome variables were included as predictors, the selected lipids were the previously-selected LysoPC(18:2), PC(32:1), PC(34:2e), TG(17:1/18:1/18:2) and TG(50:5). The cross-validated prediction performance of the “Lipids and MS” model consisting of the selected lipids and the MS variables was better than the model with the MS variables only (AUC 0.967 vs. 0.956; IDI 0.049; $p < 0.05$).

When both the FINDRISC and metabolic syndrome variables were included as predictors, the selected lipids were the same as with the “Lipids and MS” model except for the TG(50:5), which was replaced by the PC(36:1). The cross-validated prediction performance of the “Lipids, FINDRISC and MS” model consisting of the selected lipids and the FINDRISC and MS variables was better than the model with the FINDRISC and MS variables only (AUC 0.967 vs. 0.955; IDI = 0.059; $p < 0.05$).

3.5. Validation of the Predictive Model

The models fit with the discovery set were then tested by predicting the validation set. The findings with the discovery

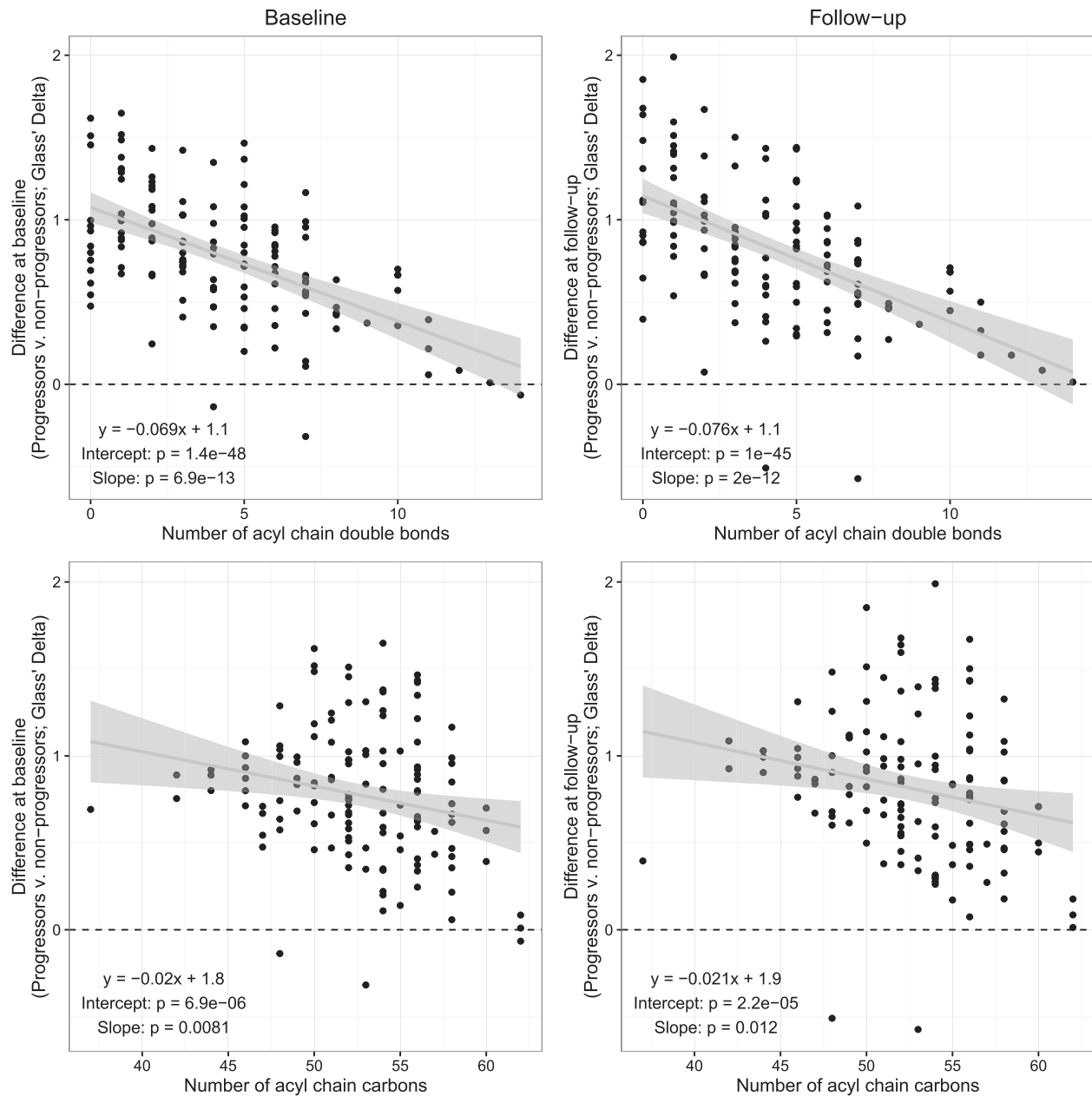


Fig. 3 – Standardized difference in triacylglycerols (TG) between progressors and non-progressors (Glass' delta effect size; y-axis) in the discovery set, shown as a function of the number of acyl chain double bonds at baseline and follow-up (top-left and top-right, respectively), and as a function of the number of acyl chain carbon atoms at baseline and follow-up (bottom-left and bottom-right, respectively).

set were replicated with the validation set (Table 4; Supplementary Fig. S4): the “Lipids” model was better than the random baseline (area under the receiver operating characteristic curve, AUC “Lipids” = 0.776 vs. AUC “Random” = 0.50; integrated discrimination improvement, IDI = 0.259; $p < 0.05$ for the test of positive IDI), the “Lipids and FINDRISC” model was better than the “FINDRISC” model (AUC 0.757 vs. 0.651; IDI 0.213; $p < 0.05$), the “Lipids and MS” model was better than the “MS” model (AUC 0.835 vs. 0.800; IDI 0.065; $p < 0.01$), and the “Lipids, FINDRISC and MS” model was better than the “FINDRISC and MS” model (AUC 0.840 vs. 0.813; IDI 0.89; $p < 0.05$).

When testing the models with the FPG-matched subset of the validation set, the improvement in the predictive power of the lipids remained significant in all comparisons while the FPG was ruled out as a confounding factor (Table 4; Supplementary Fig. S5): the “Lipids” model was better than the random baseline (AUC “Lipids” = 0.698 vs. AUC “Random” = 0.50; IDI = 0.203; $p < 0.05$), the “Lipids and FINDRISC” model was better than the “FINDRISC” model (AUC 0.692 vs. 0.561; IDI 0.178; $p < 0.05$), the “Lipids and MS” model was better than the “MS” model (AUC 0.597 vs. 0.560; IDI 0.049; $p < 0.05$), and the “Lipids, FINDRISC and MS” model was better than the “FINDRISC and MS” model (AUC 0.616 vs. 0.558; IDI 0.067; $p < 0.05$).

Table 3 – List of lipids selected to the models predicting progression to T2DM based on the baseline observations.

| Lipid name | Model | | | |
|--------------------|--------|---------------------|-------------------------|---------------|
| | Lipids | Lipids and FINDRISC | Lipids, FINDRISC and MS | Lipids and MS |
| LysoPC(18:2) | X | X | X | X |
| PC(32:1) | | | X | X |
| PC(34:2e) | | | X | X |
| PC(42:6e) | X | X | | |
| TG(17:1/18:1/18:2) | | | X | X |
| TG(50:1) | X | X | | |
| TG(54:5) | X | X | | |
| PC(36:1) | | | X | |
| TG(50:5) | | | | X |
| TG(51:5) | | X | | |
| TG(56:4) | X | | | |

Selection of a lipid (row) to a model (column) is marked by a cross ("X").

Also in the high-risk subset of the validation set (progressors and prediabetic non-progressors), the improvement in the predictive power of the lipids remained significant in all the comparisons (Table 4; Supplementary Fig. S6): the "Lipids" model was better than the random baseline (AUC "Lipids" = 0.749; AUC "Random" = 0.50; IDI = 0.244; $p < 0.05$), the "Lipids and FINDRISC" model was better than the "FINDRISC" model (AUC 0.740 vs. 0.639; IDI 0.198; $p < 0.05$), the "Lipids and MS" model was better than the "MS" model (AUC 0.802 vs. 0.766; IDI 0.067; $p < 0.05$), and the "Lipids, FINDRISC and MS" model was better than the "FINDRISC and MS" model (AUC 0.811 vs. 0.780; IDI 0.095; $p < 0.05$).

In 9 out of 11 lipids selected by the models, the direction of difference between progressors and non-progressors was

fully consistent in the discovery set at baseline and at follow-up, in the validation set, as well as in the FPG-matched subsets of the discovery and validation sets (Fig. 4). For four of the lipids—PC(32:1), PC(36:1), TG(50:1) and TG(50:5)—the difference between progressors and non-progressors was strongest in the matched normal-FPG subset of the discovery set. Also for PC(34:2e) and PC(42:6e), the difference was stronger in the normal FPG-subset of the discovery set than the entire discovery set. This finding gives indication that the identified lipidomic pattern is present across the stages of the development of T2DM. Especially changes in the six aforementioned lipids may be indicative of metabolic changes that eventually lead to T2DM.

Table 4 – Prediction performance of the models (rows) in the following data sets (columns): discovery set (cross-validated), validation set as well as the high-risk and FPG-matched subsets of the validation set.

a) Area under receiver operating characteristic (AUC) curve

| Model | Data set | | | |
|-------------------------|----------------------|----------------------|-----------------------|-------------------------|
| | Discovery | Validation, all | Validation, high-risk | Validation, FPG-matched |
| Lipids, FINDRISC and MS | 0.967 (0.944, 0.974) | 0.840 (0.823, 0.854) | 0.811 (0.787, 0.824) | 0.616 (0.561, 0.667) |
| Lipids and MS | 0.967 (0.960, 0.974) | 0.835 (0.814, 0.853) | 0.802 (0.779, 0.821) | 0.597 (0.384, 0.645) |
| FINDRISC and MS | 0.955 (0.945, 0.96) | 0.813 (0.796, 0.828) | 0.780 (0.756, 0.793) | 0.558 (0.444, 0.592) |
| MS | 0.956 (0.949, 0.959) | 0.800 (0.789, 0.810) | 0.766 (0.748, 0.775) | 0.56 (0.444, 0.582) |
| Lipids | 0.842 (0.829, 0.851) | 0.776 (0.757, 0.785) | 0.749 (0.744, 0.769) | 0.698 (0.605, 0.781) |
| Lipids and FINDRISC | 0.864 (0.852, 0.874) | 0.757 (0.719, 0.777) | 0.740 (0.710, 0.763) | 0.692 (0.594, 0.779) |
| FINDRISC | 0.755 (0.755, 0.755) | 0.651 (0.651, 0.651) | 0.639 (0.639, 0.639) | 0.561 (0.469, 0.660) |
| Random | 0.500 (0.500, 0.500) | 0.500 (0.500, 0.500) | 0.500 (0.500, 0.500) | 0.500 (0.500, 0.500) |

b) Integrated discrimination improvement (IDI) of a lipid model compared to the respective non-lipid model

| Model | Model | | | |
|-------------------------|----------------------|----------------------|-----------------------|-------------------------|
| | Discovery | Validation, all | Validation, high-risk | Validation, FPG-matched |
| Lipids | 0.359 (0.344, 0.368) | 0.259 (0.218, 0.304) | 0.244 (0.205, 0.288) | 0.203 (0.118, 0.279) |
| Lipids and FINDRISC | 0.282 (0.266, 0.296) | 0.213 (0.175, 0.248) | 0.198 (0.164, 0.231) | 0.178 (0.073, 0.273) |
| Lipids and MS | 0.059 (0.041, 0.075) | 0.089 (0.003, 0.165) | 0.095 (−0.002, 0.177) | 0.067 (−0.001, 0.141) |
| Lipids, FINDRISC and MS | 0.049 (0.033, 0.061) | 0.065 (0.023, 0.120) | 0.067 (0.016, 0.127) | 0.049 (0.001, 0.099) |

Models are assessed by area under receiver operating characteristic curve (a). Lipid marker models are compared to their respective non-lipid counterparts by integrated discrimination improvement (IDI; b). Positive IDI means improvement; all lipid models were better than their non-lipid counterpart model in each data set with $p < 0.05$.

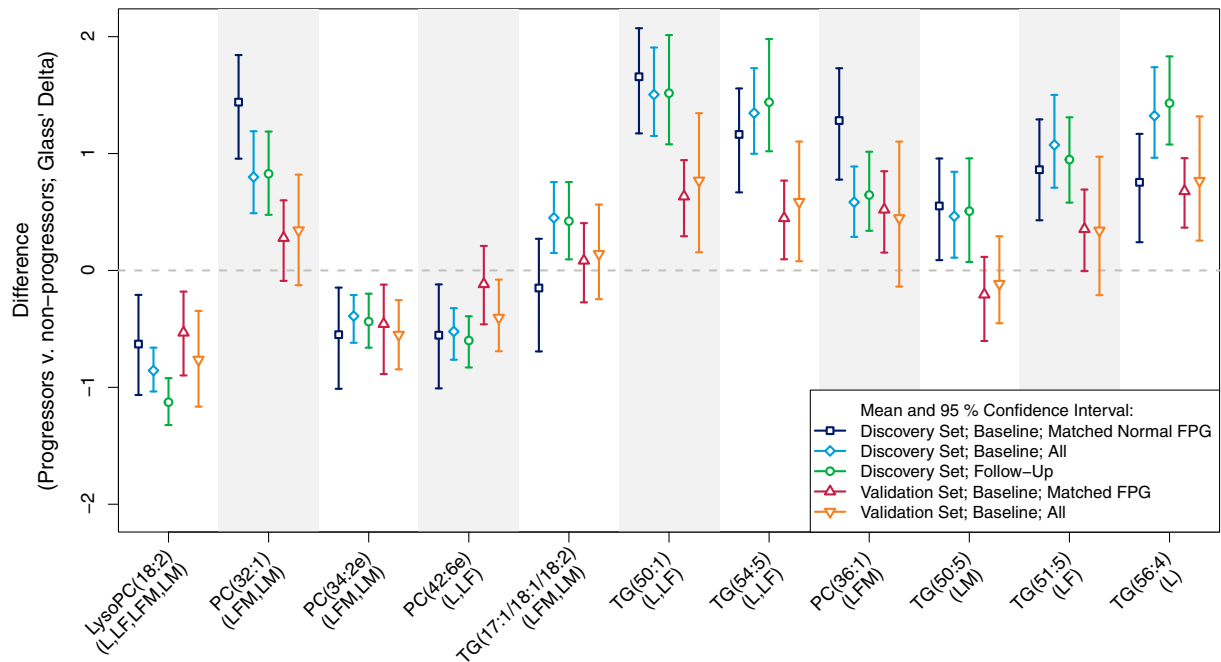


Fig. 4 – Standardized differences of the selected marker lipids between progressors and non-progressors (Glass' delta; y-axis) in the data sets of the study: the matched normal-FPG subset of the discovery group at baseline (dark blue squares), the entire discovery set (light blue diamonds), the discovery set at five-year follow-up (green circles), the FPG-matched subset of the validation set at baseline (red triangles) and the entire validation set at baseline (orange triangles). Error bars show the 95% quantiles of the difference, computed over bootstrap resamples. In progressors, five of the lipids have a change that is consistent over all the data sets: two of the selected lipids, LysoPC(18:2) and PC(34:2e), have a consistent decrease, and three triacylglycerols, TG(50:1), TG(54:5) and TG(56:4), have a consistent increase. Four lipids, PC(32:1), PC(36:1), TG(50:1) and TG(50:5) are potential early markers, since their differences between progressors and non-progressors are strongest in the matched normal-FPG subset of the discovery set. Abbreviations of the models, for which the lipid was selected, are shown on the x-axis label along with the lipid name (L: Lipids; LF: Lipids and FINDRISC; LFM: Lipids, FINDRISC and MS; LM: Lipids and MS).

4. Discussion

Risk prediction models for T2DM could be helpful for clinical decision making by identifying subjects that are likely to benefit from an early intervention. In this study, we identified a lipid molecular signature that improves the prediction of progression to T2DM compared to established risk factors. Overall, baseline levels of TGs and PCs were associated with progression to T2DM while ether PCs were inversely associated. This lipid signature persisted over the five-year follow-up period. Results with the validation cohort and its subsets indicate that the prediction models have the potential for generalizing beyond the present study.

Our findings are in agreement with several previous studies: In the KORA and EPIC-Postdam cohorts [11,13], similar associations were observed between PCs, ether PCs, LysoPCs and the risk of T2DM. LysoPC(18:2) was inversely associated with dysglycemia and T2DM risk in the RISC and Botnia cohorts, respectively [15]. Association of TGs with the progression to T2DM, particularly the TGs of low double bond content, is in agreement with findings from the Framingham Heart Study [18]. Our results are also in agreement with the lipid-based classification results from the AusDiab study [31],

where the inclusion of plasma DGs and TGs improved prediction performance compared to a model based on the total TGs and HbA1c. These studies, along with the present study, provide increasing evidence for the distinct molecular lipid profile in individuals, who are about to progress to T2DM. Furthermore, the results indicate that lipids can improve prediction performance as compared to established risk factors.

In agreement with previous studies, we have demonstrated that the distinct lipid profile in progression to T2DM can also be used for stratifying subjects within a risk group. Compared to bulk blood cholesterol and total TGs, a set of molecular lipid species may provide a more detailed view of the altered lipid homeostasis in subjects with T2DM in development. As a novel finding, we also found that the lipid profile associated with diabetes was persistent in time. To further test the utility of lipid predictors independently of FPG, we demonstrated that the models remained predictive of T2DM in an FPG-matched subset of the validation cohort.

We found that several phospholipids were associated with progression to T2DM. Phospholipids, such as PCs, ether PCs and LysoPCs, are the main components of the surface layer of lipoproteins [19]. Specifically, LysoPCs are predominantly found in the HDL fraction of healthy subjects [32,33]. It is well-known that hyperglycemia caused by T2DM leads to

alterations in lipid metabolism, including enhanced HDL clearance, decreased apoA-1 transcription and accelerated HDL glycation [34]. Lipid and structure-function studies suggest that alterations of the lipidome and the negative-charged enrichment of phospholipids in the smaller HDL sub-fraction play a crucial role in HDL dysfunctionality, particularly in T2DM. A recent study [35] demonstrated that in T2DM the HDL particles have a different surface polarity, which is associated with different lipid composition in the HDL particles. The authors of this study hypothesized that the decreased size forces CEs and TGs to emerge from the HDL core to the surface, making the outer surface of HDL more hydrophobic.

In agreement with several recent studies, we found that LysoPC(18:2) was decreased in subjects that progressed to T2DM [11,15,36]. LysoPCs are formed either by phospholipase A2-induced hydrolysis or by oxidation of PCs in phospholipid-containing structures, such as LDL and cell membranes. The decreased levels of LysoPC(18:2) observed in this study could result either (1) from the decreased activity of PLA2 [37], (2) from the increased breakdown or enhanced clearance of the lipid from circulation by metabolically active tissues, or (3) from the influence of dietary factors. In the present study, the total LysoPC levels were not decreased, rendering it unlikely that the PLA2 activity would be decreased. On the other hand, a recent study [36] found that nutrition can influence circulating LysoPC levels: broad decreases in circulating LysoPCs were observed in fat-fed mice while changes of few individual LysoPCs were observed in the liver, muscle and adipose tissues. In agreement with this explanation, in our data, LysoPC(18:2) was correlated with its precursor PCs that are associated with fat intake [38].

In a more specific T2DM risk stratification study [39], LysoPC(18:2) was reported as a selective marker of isolated impaired glucose tolerance, which is a high-risk state of T2DM. LysoPC(18:2) did not predict isolated impaired fasting glucose in that study. On the other hand, decreased levels of LysoPC(18:2) have also been reported to be inversely associated with obesity [40] as well as with BMI and age [41]. Taken together, while decreased levels of LysoPC(18:2) have been found to be associated with increased risk of T2DM in several studies, including ours, more studies are needed in order to clarify the mechanisms behind this association.

In agreement with previous studies [13,42], we also observed an inverse association between T2DM and ether-linked PCs. This may be explained at least in part by the enrichment of ether phospholipids in the HDL fraction [19,32]. T2DM has been found to be associated with both increased oxidative stress and inflammation. The concentrations of PE plasmalogens, for instance, were earlier found [43] to be decreased in T2DM subjects while hydroxylated FAs were increased. Among the selected predictor lipids, PC(42:6e) was down-regulated in progressors. A more unsaturated PC with the same carbon number has recently been reported [44] as down-regulated in another high-risk group—American Indians—who developed T2DM after a similar 5.5 year follow-up. The other two selected predictive phosphatidylcholines—PC(32:1) and PC(34:2e)—were also reported [11] with up-regulation and down-regulation, respectively, in progressors of the KORA cohort.

TGs with low double bond content are predictive of NAFLD [17] and insulin resistance [19]. The association of these TGs with

the progression of T2DM may, thus, reflect increased *de novo* lipogenesis, which is increased in NAFLD [45,46]. Our study supports the hypothesis that lipotoxicity, associated with *de novo* lipogenesis in the liver [46] and accompanied by insulin resistance, might in part be reflected by alterations in specific circulating lipid species in plasma. The lipid profiles observed in progressors to T2DM may therefore, in part, reflect the contribution of fatty liver to the progression of T2DM.

Distribution of FPG levels in the discovery study was markedly different between the progressors and non-progressors, which is a limitation of this study. However, the FPG distributions of the two groups were wider and more overlapping in the validation set while the lipid models remained predictive of progression. In spite of the aforementioned limitation, we presented an analysis of the subset of the validation set, where the effect of FPG as a confounding factor was eliminated, leaving only lipid models predictive of the progression. In the matched normal-FPG subset of the discovery set, we observed similar or even pronounced effect sizes compared to the entire discovery set. Since the total number of T2DM progressors was rather small in the validation set, it is clear that the findings need to be replicated in future studies in addition to the validation presented here.

Overall, gender differences in the risk, pathophysiology as well as complications of T2DM have been reported [47]. Many factors, including nutrition, lifestyle, and hormonal differences have been proposed as an explanation. These factors have some influence on metabolism as well as on levels of blood lipids and metabolites [48,49]. In T2DM subjects, the risk of coronary heart disease (CHD) has been observed [50] to increase more in women than in men although the incidence of CHD in non-diabetics is lower in women than in men. The greater effect of dyslipidemia on the risk of CHD in women compared to men is hypothesized to be one of the reasons. This hypothesis could not be tested in this males-only study and it remains an interesting avenue for future research.

The METSIM sample analyzed in the present study is focused on males, which is both a strength and limitation of the study. Gender differences can be ruled out as potential confounding factors in this study setting. On the other hand, the findings cannot be directly generalized across genders. Some gender differences in the plasma lipidome are known from earlier studies [41,51]: Elevated levels of CEs, TGs, ceramides and LysoPCs have been reported [41] in males compared to females while the levels of SMs and PSEs have been found to be lower in males. Gender differences in SMs were also reported in another study [52]. In the females-only DIWA study [53], most TG species—including TG(54:5) and TG(50:1)—were up-regulated in T2DM, as we observed in our males-only study. Also several PCs, including PC(36:1) and PC(32:1), showed agreeing changes although they did not reach the level of significance.

There is recent evidence that 1-deoxysphingolipids are predictive of T2DM progression [54]. These lipids were previously suggested to be associated with diabetic neuropathy but their elevation is seen also before the diagnosis [55]. Interestingly, 1-deoxysphinganine has been demonstrated to be cytotoxic for pancreatic beta-cells [56]. In our study, SMs and ceramides were not included in the predictive models. However, the lipidomics panel used did not include the deoxy sphingolipids.

Given the increasing prevalence of T2DM, there is a demand for biomarkers that change at a relatively late stage of progression to T2DM. These biomarkers could make it possible to identify subjects at the greatest risk of developing the disease. Specific metabolite biomarkers derived from metabolomic and lipidomic studies can already now be cost-effectively measured at the clinic [57]. Taken together, our study adds more evidence to the hypothesis that plasma lipids can predict progression to T2DM. We also show that the lipid signature of T2DM is persistent over time. The complementary nature of the FPG and lipid predictors suggests that elevated FPG and lipids may in part reflect different underlying pathologies that result in T2DM. Our findings may therefore pave the way towards a diagnostic application with the T2DM risk-assessment supported by a panel of lipids measured routinely at the clinic.

Author Contributions

The author contributions are as follows: study concept and design (JN, JK, MO); acquisition of data (PP, TH); analysis and interpretation of data (TS, IB-P, LY, MO); drafting first version of the manuscript (LY, IB-P, TS); critical revision of the manuscript (all authors).

Funding

This work was supported by the EU FP7 programme project DEXLIFE (Grant agreement # 279228). IB-P is grateful to the Novo Nordisk Foundation for her Clinical Research Fellow grant.

Acknowledgments

The authors would like to thank Leena Öhrnberg and Ulla Lahtinen (VTT Technical Research Centre of Finland, Espoo, Finland) for technical assistance in lipidomics analysis.

Disclosure Statement

All authors declare that there is no duality of interest associated with their contribution to this manuscript.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.metabol.2017.08.014>.

REFERENCES

- [1] Porte Jr D, Kahn SE. Beta-cell dysfunction and failure in type 2 diabetes: potential mechanisms. *Diabetes* 2001;50(Suppl. 1):S160–3.
- [2] Tabak AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. *Lancet* 2012;379:2279–90.
- [3] Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 2001;344:1343–50.
- [4] Faerch K, Hulman A, Solomon TP. Heterogeneity of pre-diabetes and type 2 diabetes: implications for prediction, prevention and treatment responsiveness. *Curr Diabetes Rev* 2015.
- [5] Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003;26: 725–31.
- [6] Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
- [7] Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011;343:d7163.
- [8] Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* 2010;5:e13953.
- [9] Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17:448–53.
- [10] Stancakova A, Civelek M, Saleem NK, Soininen P, Kangas AJ, Cederberg H, et al. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* 2012;61:1895–902.
- [11] Wang-Sattler R, Yu Z, Herder C, Messias AC, Floegel A, He Y, et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol Syst Biol* 2012;8:615.
- [12] Gall WE, Beebe K, Lawton KA, Adam KP, Mitchell MW, Nakhle PJ, et al. Alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One* 2010;5:e10883.
- [13] Floegel A, Stefan N, Yu Z, Muhlenbruch K, Drogan D, Joost HG, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 2013;62:639–48.
- [14] Cobb J, Eckhart A, Perichon R, Wulff J, Mitchell M, Adam KP, et al. A novel test for IGT utilizing metabolite markers of glucose tolerance. *J Diabetes Sci Technol* 2015;9:69–76.
- [15] Ferrannini E, Natali A, Camastra S, Nannipieri M, Mari A, Adam KP, et al. Early metabolic markers of the development of dysglycemia and type 2 diabetes and their physiological significance. *Diabetes* 2013;62:1730–7.
- [16] Alshehry ZH, Mundra PA, Barlow CK, Mellett NA, Wong G, McConville MJ, et al. Plasma lipidomic profiles improve upon traditional risk factors for the prediction of cardiovascular events in type 2 diabetes. *Circulation* 2016.
- [17] Oresic M, Hyotylainen T, Kotronen A, Gopalacharyulu P, Nygren H, Arola J, et al. Prediction of non-alcoholic fatty-liver disease and liver fat content by serum molecular lipids. *Diabetologia* 2013;56:2266–74.
- [18] Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest* 2011;121:1402–11.
- [19] Kotronen A, Velagapudi VR, Yetukuri L, Westerbacka J, Bergholm R, Ekroos K, et al. Saturated fatty acids containing triacylglycerols are better markers of insulin resistance than total serum triacylglycerol concentrations. *Diabetologia* 2009; 52:684–90.
- [20] Haus JM, Kashyap SR, Kasumov T, Zhang R, Kelly KR, Defronzo RA, et al. Plasma ceramides are elevated in obese subjects with type 2 diabetes and correlate with the severity of insulin resistance. *Diabetes* 2009;58:337–43.
- [21] Boon J, Hoy AJ, Stark R, Brown RD, Meex RC, Henstridge DC, et al. Ceramides contained in LDL are elevated in type 2

- diabetes and promote inflammation and skeletal muscle insulin resistance. *Diabetes* 2013;62:401–10.
- [22] Adams LA, Waters OR, Knudman MW, Elliott RR, Olynyk JK. NAFLD as a risk factor for the development of diabetes and the metabolic syndrome: an eleven-year follow-up study. *Am J Gastroenterol* 2009;104:861–7.
- [23] Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* 2009;58:1212–21.
- [24] Nygren H, Seppanen-Laakso T, Castillo S, Hyotylainen T, Oresic M. Liquid chromatography-mass spectrometry (LC-MS)-based lipidomics for studies of body fluids and tissues. *Methods Mol Biol* 2011;708:247–57.
- [25] Fraley C, Raftery AE. Model-based methods of classification: using the mclust software in chemometrics. *J Stat Soft* 2007; 18:1–13.
- [26] Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: linear and nonlinear mixed effects models. R Package Version 3.1103 ed. ; 2014.
- [27] Efron B, Gong G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am Stat* 1983;37:36–48.
- [28] Harrell F. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer International Publishing; 2015.
- [29] Davison AC, Hinkley DV. Bootstrap Methods and their Application. Cambridge University Press; 2013.
- [30] Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.
- [31] Wong G, Barlow CK, Weir JM, Jowett JB, Magliano DJ, Zimmet P, et al. Inclusion of plasma lipid species improves classification of individuals at risk of type 2 diabetes. *PLoS One* 2013; 8:e76577.
- [32] Wiesner P, Leidl K, Boettcher A, Schmitz G, Liebisch G. Lipid profiling of FPLC-separated lipoprotein fractions by electrospray ionization tandem mass spectrometry. *J Lipid Res* 2009;50:574–85.
- [33] Yetukuri L, Soderlund S, Koivuniemi A, Seppanen-Laakso T, Niemela PS, Hyvonen M, et al. Composition and lipid spatial distribution of HDL particles in subjects with low and high HDL-cholesterol. *J Lipid Res* 2010;51:2341–51.
- [34] Drew BG, Rye KA, Duffy SJ, Barter P, Kingwell BA. The emerging role of HDL in glucose metabolism. *Nat Rev Endocrinol* 2012;8:237–45.
- [35] Amigo N, Mallol R, Heras M, Martinez-Hervas S, Blanco Vaca F, Escola-Gil JC, et al. Lipoprotein hydrophobic core lipids are partially excluded to surface in smaller HDL: “herniated” HDL, a common feature in diabetes. *Sci Rep* 2016;6:19249.
- [36] Barber MN, Risis S, Yang C, Meikle PJ, Staples M, Febbraio MA, et al. Plasma lysophosphatidylcholine levels are reduced in obesity and type 2 diabetes. *PLoS One* 2012;7:e41456.
- [37] Goncalves I, Edsfieldt A, Ko NY, Grufman H, Berg K, Bjorkbacka H, et al. Evidence supporting a key role of Lp-PLA2-generated lysophosphatidylcholine in human atherosclerotic plaque inflammation. *Arterioscler Thromb Vasc Biol* 2012;32:1505–12.
- [38] Floegel A, von Ruesten A, Drogan D, Schulze MB, Prehn C, Adamski J, et al. Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam. *Eur J Clin Nutr* 2013;67:1100–8.
- [39] Cobb J, Eckhart A, Motsinger-Reif A, Carr B, Groop L, Ferrannini E. Alpha-hydroxybutyric acid is a selective metabolite biomarker of impaired glucose tolerance. *Diabetes Care* 2016;39:988–95.
- [40] Heimerl S, Fischer M, Baessler A, Liebisch G, Siguener A, Wallner S, et al. Alterations of plasma lysophosphatidylcholine species in obesity and weight loss. *PLoS One* 2014;9:e111348.
- [41] Weir JM, Wong G, Barlow CK, Greeve MA, Kowalczyk A, Almasy L, et al. Plasma lipid profiling in a large population-based cohort. *J Lipid Res* 2013;54:2898–908.
- [42] Meikle PJ, Wong G, Barlow CK, Weir JM, Greeve MA, Macintosh GL, et al. Plasma lipid profiling shows similar associations with prediabetes and type 2 diabetes. *PLoS One* 2013;8:e74341.
- [43] Colas R, Pruneta-Deloché V, Guichardant M, Luquain-Costaz C, Cugnet-Anceau C, Moret M, et al. Increased lipid peroxidation in LDL from type-2 diabetic patients. *Lipids* 2010;45:723–31.
- [44] Zhao J, Zhu Y, Hyun N, Zeng D, Uppal K, Tran VT, et al. Novel metabolic markers for the risk of diabetes development in American Indians. *Diabetes Care* 2015;38:220–7.
- [45] Donnelly KL, Smith CI, Schwarzenberg SJ, Jessurun J, Boldt MD, Parks EJ. Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J Clin Invest* 2005;115:1343–51.
- [46] Kotronen A, Seppanen-Laakso T, Westerbacka J, Kiviluoto T, Arola J, Ruskeepaa AL, et al. Hepatic stearoyl-CoA desaturase (SCD)-1 activity and diacylglycerol but not ceramide concentrations are increased in the nonalcoholic human fatty liver. *Diabetes* 2009;58:203–8.
- [47] Kautzky-Willer A, Harreiter J, Pacini G. Sex and gender differences in risk, pathophysiology and complications of type 2 diabetes mellitus. *Endocr Rev* 2016;37:278–316.
- [48] Uusitupa M, Hermansen K, Savolainen MJ, Schwab U, Kolehmainen M, Brader L, et al. Effects of an isocaloric healthy Nordic diet on insulin sensitivity, lipid profile and inflammation markers in metabolic syndrome — a randomized study (SYSDIET). *J Intern Med* 2013;274:52–66.
- [49] Scalbert A, Brennan L, Manach C, Andres-Lacueva C, Dragsted LO, Draper J, et al. The food metabolome: a window over dietary exposure. *Am J Clin Nutr* 2014;99:1286–308.
- [50] Juutilainen A, Kortelainen S, Lehto S, Ronnemaa T, Pyorala K, Laakso M. Gender difference in the impact of type 2 diabetes on coronary heart disease risk. *Diabetes Care* 2004;27:2898–904.
- [51] Cai X, Perttula K, Pajouh S, Hubbard A, Nomura D, Rappaport S. Untargeted lipidomic profiling of human plasma reveals differences due to race, gender and smoking status. *Metabolomics* 2014;4:1.
- [52] Ishikawa M, Maekawa K, Saito K, Senoo Y, Urata M, Murayama M, et al. Plasma and serum lipidomics of healthy white adults shows characteristic profiles by subjects' gender and age. *PLoS One* 2014;9:e91806.
- [53] Stahlman M, Fagerberg B, Adiels M, Ekroos K, Chapman JM, Kontush A, et al. Dyslipidemia, but not hyperglycemia and insulin resistance, is associated with marked alterations in the HDL lipidome in type 2 diabetic subjects in the DIWA cohort: impact on small HDL particles. *Biochim Biophys Acta* 1831;2013:1609–17.
- [54] Mwinyi J, Bostrom A, Fehrer I, Othman A, Waeber G, Marti-Soler H, et al. Plasma 1-deoxysphingolipids are early predictors of incident type 2 diabetes mellitus. *PLoS One* 2017; 12:e0175776.
- [55] Dohrm MF, Othman A, Hirshman SK, Bode H, Alecu I, Fahndrich E, et al. Elevation of plasma 1-deoxy-sphingolipids in type 2 diabetes mellitus: a susceptibility to neuropathy? *Eur J Neurol* 2015;22:806–14 [e55].
- [56] Zuellig RA, Hornemann T, Othman A, Hehl AB, Bode H, Guntert T, et al. Deoxysphingolipids, novel biomarkers for type 2 diabetes, are cytotoxic for insulin-producing cells. *Diabetes* 2014;63:1326–39.
- [57] Mamtani M, Kulkarni H, Wong G, Weir JM, Barlow CK, Dyer TD, et al. Lipidomic risk score independently and cost-effectively predicts risk of future type 2 diabetes: results from diverse cohorts. *Lipids Health Dis* 2016;15:67.